J|A|C|S
COMMUNICATIONS

Published on Web 06/18/2003

Integrated Computational and Experimental Approach for Lead Optimization and Design of Compstatin Variants with Improved Activity

John L. Klepeis,[†] Christodoulos A. Floudas,*[,†] Dimitrios Morikis,[‡] C. G. Tsokos,[§] E. Argyropoulos,[§] L. Spruce,[§] and John D. Lambris*[,§]

Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544, Department of Chemical and Environmental Engineering, University of California at Riverside, Riverside, California 92521, and Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received February 24, 2003; E-mail: floudas@titan.princeton.edu

The problem of protein design truly tests the capacity to understand the relationship between the amino acid sequence of a protein and its three-dimensional structure.[1] The problem, first suggested almost two decades ago,[2] begins with a known protein structure and requires the determination of an amino acid sequence compatible with this structure. Computational protein design allows for the screening of large sectors of sequence space, leading to the possibility of a much broader range of functional properties among the selected sequences when compared to experimental techniques. The first validated computational design of a full sequence was accomplished by using a combination of a backbone-dependent rotamer library and a dead-end elimination-based algorithm.[2a,3] Despite such breakthroughs, understanding structural−functional property relationships remains an unsolved problem. In this communication, a study of computation and experiment is presented and applied to the problem of immunological property improvement for a synthetic peptide. At the heart of the methodology lies a novel two-stage computational protein design method used not only to select and rank sequences for a particular fold but also to validate the stability and specificity of the fold for these selected sequences. The parent peptide, compstatin, a 13-residue peptide with a disulfide bridge, inhibits complement activation and has been resolved structurally via NMR.[4] The application of the presented approach has led to the identification of sequences with predicted improvements in inhibition activity, with subsequent verification of inhibitory activity using complement inhibition assays.

Compstatin (ICVVQDWGHHRCT), a candidate for being a therapeutic agent, inhibits complement component C3, a central player in the activation of all complement pathways. Unchecked complement activation causes host cell damage, which may lead to one of more than 25 pathological conditions, including autoimmune diseases, stroke, heart attack, and burn injuries.[5] Compstatin was initially identified through screening of a phage-displayed random peptide library,[6] and subsequent rational design studies indicated that Val^3 as well as the four residues of the β-turn are essential, although not sufficient, conditions for retaining activity.[4,7] In particular, the flexibility of the turn was found to be important, with more stable type-I β-turn sequences leading to lower or no activity. Compstatin also possesses a hydrophobic cluster (residues 1, 2, 3, 4, 12, and 13) that is held together by a disulfide bridge, but this component is also not sufficient for activity. The difficulty in the optimization of the compstatin system has been demonstrated through both experimental combinatorial and rational design techniques,[7] with both studies leading to the identification of only a 2-fold more active analogue.

The first stage of the proposed in silico design approach involves the selection of sequences compatible with the backbone template (from NMR-average structure of compstatin[4]) through the solution of an integer linear optimization problem (see Supporting Information). A general and well-established distance-dependent potential, with the implicit inclusion of side-chain interactions and amino acid specificity,[8] is used in the objective. In light of the results of the experimental studies for the rationally designed peptides, a directed set of computational design studies was performed, which highlights the underlying hypothesis of the approach: predicted increases in fold stability and specificity, while maintaining certain important functional components, are equivalent to real increases in functionality. In this case, the disulfide bridge was enforced, and turn residues (5−8) were fixed to be those of the parent sequence.

After designing the experiment to be consistent with those features found to be essential for compstatin activity, six residue positions were selected to be optimized. Of these six residues, positions 1, 4, and 13 belong to the hydrophobic cluster, while positions 9, 10, and 11 are between the β-turn and the C-terminal cysteine. To maintain the hydrophobic cluster, positions 1, 4, and 13 were allowed to select only from those residues defined as belonging to the hydrophobic set (A,F,I,L,M,V,Y), including threonine for position 13 (to allow for the selection of the parent peptide residue). In positions 9, 10, and 11 all residues were allowed. Using a rank-ordered list of the 50 lowest-lying energy sequences, the residues found to have more than 10% representation at each position (in order of decreasing count) were: (i) A and V at position 1; (ii) Y and V at position 4; (iii) T, F, and A at position 9; (iv) H at position 10; (v) T, V, A, F, and H at position 11; and (vi) V, A, and F at position 13. The selection of histidine at position 10 agrees with the parent peptide sequence, while position 11 is found to have the largest variation in composition. At position 9 a subset of those residues chosen for position 11 are selected. Although valine is strong at all positions in the hydrophobic cluster, the results for position 4 contrast those at positions 1 and 13 in that tyrosine, not valine, is the preferred choice for the lowest- and many other low-lying energy sequences.

On the basis of the sequence selection results, several optimal sequences were considered in the second stage of the design procedure (see Supporting Information). Fold stability and specificity validation of these sequences is based on the calculation of ensemble probabilities for a flexible compstatin template using full atom force field and deterministic global optimization.[9] The fold stability predictions were analyzed according to their relative probabilities (to the probability for the parent peptide) by grouping results into three different classes, which correspond to those sequences exhibiting the following: (class i) more than 3×; (class ii) between 0.5 and 3×; and (class iii) <0.5 the stability of the

* Corresponding authors.
[†] Princeton University.
[‡] University of California.
[§] University of Pennsylvania.

10.1021/ja034846p CCC: $25.00 © 2003 American Chemical Society

parent peptide sequence. Two control experiments (X1: H9A and X2: I1L,H9W,T13G), for which experimental results indicated modest increases in activity, were also performed. The results indicate that both sequences belong in (ii), thereby verifying the consistency of the classification scheme.

For all subsequent calculations, position 10 was set to be histidine, while position 11 was restricted to be arginine, on the basis of the focused and broad prediction results at these positions, respectively. To assess the dominant selection of tyrosine at position 4, tyrosine was assigned to position 4 for all five sequences belonging to set A (see Supporting Information for descriptions of all sequences for sets A, B, C, and D). The results for all five sequences belong in (ii), a significant implication for the proposed binding model of the compstatin−C3 complex.[7] Previous peptide analogues with alanine substitutions on both sides of the β-turn provided activity equivalent to that of the original compstatin sequence. These observations led to the hypothesis that the presence of side interactions of Val[4] in the binding and activity of compstatin was not required, corroborating the ability to make the valine-to-tyrosine substitution at position 4. Sequence A5, with a histidine-to-phenylalanine switch at position 9, demonstrated the highest relative stability in sequence set A.

To further explore the combination of position 9 substitutions with the presence of tyrosine at position 4, three additional sequences were tested computationally (set B). These constructions represent a reduction in the number of simultaneous mutations in the parent peptide sequence (compared to set A), with the common substitution of valine at position 13. Each of the three designed sequences are predicted to have significant increases in fold stability and specificity, and belong to (i).

Set C was composed of two sequences, with the difference between them being the assignment of tyrosine and valine to position 4 of sequences C1 and C2, respectively. Although a decrease in stability is predicted, there is again strong evidence for the preference of tyrosine at position 4. The overall loss in stability is most likely due to the negative net charge balance (threonine substitutions at positions 9 and 11), which validates the placement of arginine at position 11 for the other sequence sets.

The final set of sequences, D1 and D2, resembles the set of sequences B1 and B2 with threonine instead of valine at the C-terminal position. Both sequences provide significant increases in predicted fold stability and specificity and are grouped with sequences B1, B2, and B3 in (i). For sequences D1 and D2 the differences with respect to the parent peptide sequence are isolated to the residues before and after the β-turn. Both the position 4 tyrosine and position 9 phenylalanine substitutions provide enhancements in fold stability and specificity and represent unforeseen enhancements over the parent peptide.

On the basis of these predictions, a number of the designed sequences presented above were constructed and tested experimentally for their activity, and these results were then compared to those of the theoretical predictions. To this end, one or two analogues were experimentally tested per category, with designed peptide D1 shown to be the most active compstatin analogue currently available. Figure 1 shows quantitative results from the inhibition experiments in comparison to the theoretical fold stability and specificity results and demonstrates that the experimental results are in excellent qualitative agreement with predicted increases in fold stability and specificity.

The success of the approach is significant because (i) it validates the underlying hypothesis that predicted improvements in stability and specificty can lead to improved activity when conserving certain functionally important features and (ii) such a directed formulation can be seamlessly integrated into a computational method to predict analogues with enhanced immunological activities. Experimental results indicate that the most active compstatin analogues are sequences D1 and B1, as suggested by the optimization study. The common characteristics of these two sequences are the substitutions at positions 4 and 9 (flanking the β-turn) with the combination of tyrosine at position 4 and alanine at position 9 leading to an approximate 6- to 7-fold improvement over the parent peptide compstatin. This is a significant increase in activity over analogues identified by either purely rational or experimental combinatorial design techniques.

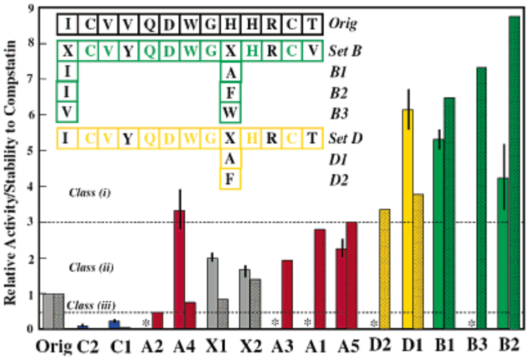

Figure 1. Comparison of relative inhibitory activity and fold stability and specificity of designed sequences. Theoretical fold stability shown as patterned bars (right), and experimental inhibitory activity shown as solid bars (left, with error bars). An asterisk (*) indicates that experimental data is not available. Three classifications of both relative fold stability and relative activity are made: class (i) more than 3 times; class (ii) between 0.5 and 3 times; and class (iii) <0.5 either the fold stability or activity of the parent compstatin sequence.

Acknowledgment. C.A.F. gratefully acknowledges support from NSF and NIH (R01 GM52032). J.D.L. gratefully acknowledges support from NIH (AI 30040 and GM 62134).

Supporting Information Available: Methodology of two-stage computational design procedure, full description of sequences tested for fold stability and specifity, and experimental data for compstatin and sequence D1 inhibition studies (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Ventura, S.; Vega, M.; Lacroix, E.; Angrand, I.; Spagnolo, L.; Serrano, L. *Nat. Struct. Biol.* **2002**, *9*, 485. (b) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425. (c) Ottesen, J. J.; Imperiali, B. *Nat. Struct. Biol.* **2001**, *8*, 535. (d) Koehl, P.; Levitt, M. *J. Mol. Biol.* **1999**, *293*, 1161. (e) Hill, R. B.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 1138. (f) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82.
(2) (a) Drexler, K. E. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 5275. (b) Pabo, C. *Nature* **1983**, *301*, 200.
(3) Desmet, J.; Maeyer, M. D.; Hazes, B.; Lasters, I. *Nature* **1992**, *356*, 539.
(4) Morikis, D.; Assa-Munt, N.; Sahu, A.; Lambris J. D. *Protein Sci.* **1998**, *7*, 619.
(5) Sahu, A.; Lambris J. D. *Immunol. Rev.* **2001**, *180*, 35.
(6) Sahu, A.; Kay, B. K.; Lambris J. D. *J. Immunol.* **1996**, *157*, 884.
(7) (a) Morikis, D.; Roy, M.; Sahu, A.; Torganis, A.; Jennings, P. A.; Tsokos, G. C.; Lambris J. D. *J. Biol. Chem.* **2002**, *277*, 14942. (b) Sahu, A.; Soulika, A. M.; Morikis, D.; Spruce, L.; Moore, W. T.; Lambris, J. D. *J. Immunol.* **2000**, *165*, 2491.
(8) Tobi, D.; Elber, R. *Proteins* **2000**, 41, 40.
(9) (a) Klepeis, J. L.; Floudas, C. A. *J. Global Optim.* **2003**, *25*, 113. (b) Klepeis, J. L.; Schafroth, H. D.; Westerberg, K. M.; Floudas, C. A. *Adv. Chem. Phys.* **2002**, *120*, 254. (c) Klepeis, J. L.; Floudas, C. A.; Morikis, D.; Lambris, J. D. *J. Comput. Chem.* **1998**, *20*, 1354.

JA034846P